

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 19:15:44 ; Search time 53 Seconds
(without alignments)

676.779 Million cell updates/sec

Title: US-09-934-455-6

Perfect score: 718

Sequence: 1 MESSNRSSNQSDDKQARF.....FEYLDKVLBELDSEERK 139

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phase.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	718	100.0	139	10	Q9LTC5
2	355.5	49.5	133	10	P91822
3	352.5	49.1	139	10	Q9LTC6
4	351	48.9	131	10	Q9LSX0
5	323	45.0	141	10	Q8H075
6	293	40.8	243	10	Q8LBU9
7	291.5	40.6	273	10	Q8AQU3
8	284.5	39.6	236	10	Q8SUX3
9	284.5	39.6	243	10	Q8VYM0
10	283.5	39.5	240	10	Q8H689
11	280.5	39.1	245	10	Q8GZB9
12	280	39.0	282	10	Q8M4Y9
13	279	38.9	166	10	Q8S6S1
14	277	38.6	303	10	Q949D4
15	276.5	38.5	277	10	P93392
16	276.5	38.5	345	10	Q949D2

17	273	38.0	244	10	Q9LND1
18	272.5	38.0	224	10	Q8LGR8
19	272	37.9	226	10	O22167
20	272	37.9	226	10	Q8L9K1
21	269.5	37.5	233	10	Q40479
22	269.5	37.5	237	10	Q9LW50
23	268	37.3	203	10	Q9LDB6
24	267	37.2	218	10	Q9ZNR2
25	265	36.9	291	10	Q40478
26	262	36.5	281	10	O80387
27	262	36.5	329	10	Q9LVT6
28	261.5	36.4	218	10	Q8LDC8
29	261.5	36.4	221	10	Q8VY90
30	261.5	36.4	221	10	Q8LDO4
31	261.5	36.4	250	10	Q9LU55
32	261.5	36.4	603	10	O23591
33	260	36.2	234	10	O04680
34	260	36.2	236	10	Q40476
35	260	36.2	282	10	Q9LW48
36	259.5	36.1	213	10	Q943F1
37	259.5	36.1	282	10	Q8VZ91
38	259	36.1	237	10	Q9FR02
39	258.5	36.0	202	10	Q8LXR3
40	258	35.9	201	10	Q9FKG2
41	258	35.9	210	10	Q9LEM6
42	257.5	35.9	214	10	Q9FR33
43	256.5	35.7	201	10	Q8LDL4
44	256	35.7	241	10	Q8L909
45	255	35.5	241	10	Q9C5M2

ALIGNMENTS

RESULT 1

Q9LTC5 PRELIMINARY; PRT; 139 AA.
AC Q9LTC5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Nicotiana ERBBP-3-like protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence features of the regions of 4,504,864 bp covered by sixty P1 and TAC clones.";
RT DNA Res. 7:131-135(2000).
DR EMBL; AB025608; BAA95736.1; -.
DR HSSP; O80337; 2GCC.
DR InterPro; IPR001471; TF_ERF.
DR Pfam; PF00847; AP2-Domain; 1.
DR PRINTS; PR00367; ETRSPLEENMT.
DR ProDom; PD001423; TF_ERF; 1.
DR SMART; SM00380; AP2; 1.
SQ SEQUENCE 139 AA; 16096 MW; 5B43AB05E93B4050 CRC64;

Query Match 100.0%; Score 718; DB 10; Length 139;
Best Local Similarity 100.0%; Pred. No. 3.9e-69;
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```
DR InterPro: IPR001471; TF_ERF.
DR Pfam: PF00847; AP2-Domain; 1.
DR PRINTS: PR00367; ETHRSPELEMT.
DR ProDom: PD001423; TF_ERF; 1.
DR SMART: SM00380; AP2; 1.
SQ SEQUENCE 131 AA; 14343 MW; ABD1284C83167F61 CRC64;

Query Match 48.9%; Score 351; DB 10; Length 131;
Best Local Similarity 55.7%; Pred. No. 2.6e-29;
Matches 68; Conservative 24; Mismatches 22; Indels 8; Gaps 2;

Qy 16 KOAREGVRRRPWGKFAAEIRDPNSRGARLWLGTFETAEAAAYDRAAFNLRGHILN 75
Db 11 EHGKRGVRRPWGKFAAEIRDSRKHGKRVMLGTFETAEAAAYDQAYSNRGQAALN 70
Qy 76 FPNYYPRMDYSLRPPYASSSSSSSGSTSTNVSRQNRQFEFFEYLDKVLLELLDSE 135
Db 71 FPHY-----YNGSGVSSSTAMAGSSSSASAS-SSSRQVFEFFEYLDKVLLELLBEG 122
Qy 136 ER 137
Db 123 EK 124

RESULT 5
Qy Q8H075 PRELIMINARY; PRT; 141 AA.
AC Q8H075;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN OSJNB0050N02.16.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Wing R.A., Yu Y., Soderlund C., Kim H.-R., Rambo T., Saski C.,
RA Currie J., Collura K.;
RT "Rice Genomic Sequence.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC105734; AAN87744.1; -.
KW Hypothetical protein.
SQ SEQUENCE 141 AA; 15269 MW; 802E33D920C6F51B CRC64;

Query Match 45.0%; Score 323; DB 10; Length 141;
Best Local Similarity 50.4%; Pred. No. 2.6e-26;
Matches 69; Conservative 19; Mismatches 25; Indels 24; Gaps 4;

Qy 13 QDDK-----QARFGVRRPWGKFAAEIRDPNSRGARLWLGTFETAEAAAYDRAAFN 66
Db 2 EDDKKEGSSSSYGVKRWPKFAAEIRDPNSRGARLWLGTFETAEAAAYDRAAFN 61
Qy 67 LRGLAILNFPNEY-PRM-----DDYSLRPPYASSSSSSSGSTSTNVSRQNR 115
Db 62 MGATAMLNFPDHHGAASRMTSTGSSSSSTPTPPANSAAAGGSGDRTT----- 114
Qy 116 EVFEFFEYLDKVLLELL 132
Db 115 DKVELECLDDKVLLELL 131

RESULT 6
Qy Q8LBU9 PRELIMINARY; PRT; 243 AA.
AC Q8LBU9;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Ethylene responsive element binding factor 2 (ATERF2).

OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome
annotation.";
RL Genome Biol. 0:0-0(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY086983; AAM64544.1; -.
DR InterPro: IPR001471; TF_ERF.
DR Pfam: PF00847; AP2-Domain; 1.
DR PRINTS: PR00367; ETHRSPELEMT.
DR ProDom: PD001423; TF_ERF; 1.
DR SMART: SM00380; AP2; 1.
SQ SEQUENCE 243 AA; 26728 MW; A2D4C978560CB490 CRC64;

Query Match 40.8%; Score 293; DB 10; Length 243;
Best Local Similarity 51.1%; Pred. No. 7.9e-23;
Matches 68; Conservative 19; Mismatches 30; Indels 16; Gaps 4;

Qy 16 KOAREGVRRRPWGKFAAEIRDPNSRGARLWLGTFETAEAAAYDRAAFNLRGHILN 75
Db 113 KAKHYGVRQRPWGKFAAEIRDPKNGARVWLGTFTAEADALAYDIAAFMRGSSALIN 172
Qy 76 FP---NEYPRMDYSLRPPYASSSSSSSGSTSTN---VSRQNRQFEFFEYI----- 123
Db 173 FPLRVNSGEP--DPVRIITSKRSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 230
Qy 124 --DDKVLLELLDS 134
Db 231 VGDETRVDELLVS 243

RESULT 7
Qy Q9AQU3 PRELIMINARY; PRT; 273 AA.
AC Q9AQU3;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Putative ethylene-responsive element binding protein (Putative
DE ethylene-responsive element binding protein).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Yuan Q., Ouyang S., Moffat K.S., Hill J.N., Burt P.C.,
RA Hsiao J., Zismann V., Pai G., Bowman C.L., Fujii C.Y., VanAken S.E.,
RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,
RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RT "Oryza sativa chromosome 10 BAC OSJNB0089A17 genomic sequence.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Jenkins C.N., Hsiao J.,
RA Zismann V., Pai G., Bowman C.L., Fujii C.Y., VanAken S.E., Craven B.,
RA Khalak H., Feldblyum T.V., Quackenbush J., White O., Salzberg S.L.,
RA Fraser C.M.;
```



```
QY          99 SSSSGSTS-----TNVSRQNQRVEFEFYLDDKVLEELLDS   134
      ||||| :|           :|       | | | : | | | |
Db         190 SSSNSSSCSSSSTSTSSSSSKOSVVKQESGTLVFEDLGAEYLEOLLMS   241
```

RESULT 10

Q8H6S9	PRELIMINARY;	PRT; 240 AA.
AD	Q8H6S9	
IC	Q8H6S9;	
DT	01-MAR-2003 (TRENBLurel. 23, Created)	
DT	01-MAR-2003 (TRENBLurel. 23, Last sequence update)	
DT	01-MAR-2003 (TRENBLurel. 23, Last annotation update)	
DE	Transcription factor TSRF1.	
DE	TSRF1.	
GN	Lycopersicon esculentum (Tomato).	
OS	Lycopersicon esculentum (Tomato).	
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;	
OC	Asteridae; lamiales; Solanales; Solanaceae; Solanum.	
OX	NCBI_TaxID=4081;	
OX	[1]	
RA	SEQUENCE FROM N.A.	
RP	Zhang H., Chen J., Wang X., Huang D., Huang R.;	
RP	"A tomato transcription factor regulating expression of stress	
RT	responsive genes.";	
RT	Submitted (MAR-2002) to the EMBL/GenBank/DDJ databases.	
RL	EMBL; AF494201; AAN32899.1; "	
SR	SEQUENCE 240 AA; 27031 MW; 11B2DBGA28000353 CRC64;	
DQ		

Query Match 39.5%; Score 283.5; DB 10; Length 240;
Best Local Similarity 45.7%; Pred. No. 7.9e-22;
Matches 63; Conservative 18; Mismatches 38; Indels 19;

Qy	16	QOARFPGVRRRPMWGPKAAEIRDPSSNGARLWLGTFTAEZAARAYDRAAFNLGHLALIN	75
Db	87	KEKSYGVRRRPMWGPKAAEIRDPSTENGVRVWLGTGFSADAAALYDQAASFMRGNGALIN	146
Qy	76	FNFEYYP-----RMD-----YSLAPPVASSSSSSSGSTSTNYSRONORE	116
Db	147	FPVETVRDSLDMKCHVDDCSPPVALKKRHSMKGRSTNKKVNSISKVVRVKMEVNN	206

Qy 117 VFEFYLDDKVLEELDS 134
| | | | | | | | | |
Db 207 VVFEDLGADYLEQLLSS 224

RESULT 11

Accession ID	Q8GZE9	PRELIMINARY;	PRT; 245 AA.		
QC	Q8GZE9;				
DT	01-MAR-2003	(TrenMBLrel. 23, Created)			
DT	01-WAR-2003	(TrenMBLrel. 23, Last sequence update)			
DT	01-MAR-2003	(TrenMBLrel. 23, Last annotation update)			
DE	Ethylene responsive element binding protein.				
GN	EREB.				
OS	Eucalyptus esculentum (Tomato) .				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Eukaryota; Magnoliopsida; Asteridae; lamiales; Solanales; Solanaceae; Solanum.				
CC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; lamiales; Solanales; Solanaceae; Solanum.				
OX	NCBI_TaxID=4081;				
RN	[1]	SEQUENCE FROM N.A.			
RP	STRAIN=cv. ZhongShuLiChun;				
RA	Cheng X.G., Hou Y.X.,				
RL	Submitted (APR-2002) to the EMBL/GenBank/DDSI databases.				
DR	EMBL; AF502085; AAN77067.1 ;				
SQ	SEQUENCE 245 AA; 27602 MW; 6D693744E00C2264 CRC64;				

Query Match 39.1%; Score 280.5; DB 10; Length 245;
Best Local Similarity 44.9%; Pred. NO. 1.7e-21;
Matches 62; Conservative 19; Mismatches 38; Indels 19;

QY 16 KOARFGRVRRPWPWKFAAEIRDPSRNGARLWLGTFETAEEAARAYDRAAFNLRGHLAIN 75

Db	87	KEKSYGVRRRPWGFAAEIRDSTNGVRVWLGTFDSAEDAAALAYDQAAFSMRGNSAILN	146
Qy	76	FNEYYVP-----RMD-----YSLRPPYASSSSSSSSSGSTSTNYSRQORE	116
Db	147	FPVEIVRSLRDMKCHVDCCSPVALKGRHSRKSTNSKKVNSISKVREYKVMENVNN	206
Qy	117	VFEFEYLDKVLLELLDS	134
Db	207	VVVFEDLGADYLEQLLST	224

RESULT 12

Q9M4Y9	.
ID	Q9M4Y9 PRELIMINARY; PRT; 282 AA.
AC	Q9M4Y9;
DT	01-OCT-2000 (TrEMBLrel. 15, Created)
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE	AP2-related transcription factor.
GN	CDBP.
OS	Mesembryanthemum crystallinum (Common ice plant).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC	Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.
OX	NCBI_TaxId=3544;
RN	[1]
RP	SEQUENCE FROM N.A.

NC 13002-lead, Baur B.;
RA Scharte J., Baur B.;
RT "A stress induced transcription factor of the AP2 gene family from the
RT inducible CAM-plant *Mesembryanthemum crystallinum* L.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL: AF245119; AF63205.1; -
DR HSSP: O80337; 2GCC.
DR InterPro: IPR001471; TF_ERF.
DR Pfam: PF00847; AP2-domain; 1.
DR PRINTS: PR00367; ETRSPLEMTNT.
DR ProDom: PD001423; TF_ERF; 1.
DR SMART: SM00380; AP2; 1.
SO SEQUENCE 282 AA, 30011 MW; 617C32009C535B29 CRC64;

Query Match	39.0%	Score 280;	DB 10;
Best Local Similarity	58.7%;	Pred. No. 2.3e-21;	Length 282;
Matches	61: Conservative	11: Mismatch	22: Indels
	10: Gaps		

Qy	20	FRGVRRRPWGKFAAEIRDPDSRNGARLWLGTFETAAEAARYDAYDRAAFNLRLGHAILNPP--77
	:	:
Db	138	YRGVRRRPWGKFAAEIRDPAKNGARVWLGTFTEDDAALAYDRAAFPMMRGSKALLNPPLR197

Qy		78 -NEYYP---RMDYSLRPYPYASSSSSSSGSTSNVSRQNQREV 117
	: :	
	: :	
Db	198 VNSGEPDPVRITSKRSSPERSVSSSSSESASP----KRRKKKEEV 237	

RESULT 13

Q8S6S1	PRELIMINARY;	PRT;	166 AA.
ID	Q8S6S1		
AC	Q8S6S1		
DT	01-JUN-2002 (T-EMBLrel. 21, Created)		
DT	01-JUN-2002 (T-EMBLrel. 21, Last sequence update)		
DT	01-OCT-2002 (T-EMBLrel. 22, Last annotation update)		
DE	Putative DNA binding protein.		
GN	OSUNBA003P04.20.		
OS	Oryza sativa (japonica cultivar-group).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;		
OC	Ehrhartoideae; Oryzaceae; Oryza.		
OX	NCBI_TaxID=39947;		
OX	[1]		
RN	SEQUENCE FROM N.A.		
RP	STRAIN=cv. Nipponbare;		
RC	Buell C.R.; Yuan Q.; Ouyang S.; Liu J.; Moffat K.S.; Hall J.N.;		
RA			

```

Db      209 IGSEIAAAAAAAAAAGNKRYPDPASSGSSSPSSSSSSSSSGSPKRRKRG 261
                                         ||||| | : | : | : | : |
RESULT 15
P93392          PRELIMINARY;           PRT;       277 AA.
ID   P93392;
AC   AC
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   S25-XPI DNA binding protein.
OS   Nicotiana tabacum (Common tobacco).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; lamiids; Solanales; Solanaceae; Nicotiana.
OX   NCB_I_TaxID=4097;
RN   [1]
RS   SEQUENCE FROM N.A.
RA   Xu P., Ling J., Li D., Hasegawa P.M., Bressan R.A.;
RL   Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.
DR   EMBL; U81157; AAB38748.1; -.
DR   HSSP; O80337; 2GCC.
DR   InterPro; IPR001471; TF_ERF.
DR   Pfam; PF00847; AP2-domain; 1.
DR   PRINTS; PR00367; ETHRSPLEMT.
DR   ProDom; PD001423; TF_ERF; 1.
DR   SMART; SM00380; AP2; 1.
SQ   SEQUENCE 277 AA; 31616 MW;  BDB646560199CD96 CRC64;
      Query Match              38.5%; Score 276.5; DB 10; Length 277;
      Best Local Similarity    44.9%; Pred. No. 5.2e-21;
      Matches 71; Conservative 22; Mismatches 38; Indels 27; Gaps 5;
QY   2 ESSNSSNNQSODDKOAR---FRGVRRPWGKFAAEIRDPSNRNGARLMGLTETAEAAAR 58
Db      71 EYSSKSVKIETEERPAKEKSPRGVRRPWGKFAAEIRDSIRNGVRVWLGTSDSPAAL 130
      ||| : | : | : | : | : | : | : | : | : | : | : | : | : |
QY   59 AYDRAAFNLRGHLATILNPNEYYP-----RMDD-----YSLRPYASSSSS 99
Db      131 AYDQAFLMRGTSATILNPFVETVQESLRDMKCHVDECSVVVALKKRHSLRK--KSLSK 188
      ||| : | : | : | : | : | : | : | : | : | : | : | : | : |
QY   100 SSSGSTSTNVSRQNORE---VFPEFYLDKVLLEEILDS 134
      | : : | : | : | : | : | : | : | : | : | : | : | : |
Db      189 KNSSSNSKVYREVKNVNWVWFEDLGFDYLEQLLS 226
      | : : | : | : | : | : | : | : | : | : | : | : | : |

Search completed: December 17, 2003, 19:23:11
Job time : 57 secs

```

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	355.5	49.5	133	2	D86175	hypothetical prote	
2	293	40.8	243	2	T51989	ethylene responsiv	
3	284.5	39.6	236	2	B84718	hypothetical prote	
4	276.5	38.5	277	2	T03927	DNA binding protei	
5	273	38.0	244	2	B86197	hypothetical prote	
6	272.5	38.0	225	2	A85196	EREP-2 protein (I	
7	272.5	38.0	266	2	T51988	ethylene responsiv	
8	272	37.9	161	2	T07689	transcription fact	
9	272	37.9	226	2	T00409	ethylene-responsiv	
10	269.5	37.5	233	2	T02590	DNA binding protei	
11	265	36.9	291	2	T02434	DNA binding protei	
12	262	36.5	281	2	T52189	ethylene responsiv	
13	262	36.5	300	2	T52020	ethylene responsiv	
14	261.5	36.4	603	2	E71444	probable EREP-4 -	
15	260	36.2	234	2	T07686	transcription fact	
16	260	36.2	236	2	T02432	ethylene-responsiv	
17	254	35.4	218	2	F84748	probable AP2 domai	
18	247.5	34.5	207	2	T49897	transcription fact	
19	245.5	34.2	358	2	T86579	hypothetical prote	
20	244.5	34.1	225	2	T52011	ethylene responsiv	
21	243.5	33.9	268	2	T04787	hypothetical prote	
22	239.5	33.4	222	2	T52019	ethylene responsiv	
23	238	33.1	171	2	T00432	probable AP2 domai	
24	238	33.1	298	2	T07784	AP2 domain protein	
25	237.5	33.1	212	2	T48580	hypothetical prote	
26	235.5	32.8	204	2	S49031	cadmium-induced pr	
27	235	32.7	303	2	T04541	hypothetical prote	
28	234	32.6	315	2	T47955	hypothetical prote	
29	228	31.8	262	2	E96747	hypothetical prote	

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A85001; MUID:20083488; PMID:10617198

A:Accession: A85196

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-225 <STO>

A:Cross-references: GB:NC_001268; NID:95281024; PIDN:CAB45963.1; GSPDB:GN00140

C:Genetics:

A:Gene: d14785w

A:Map position: 4

Query Match 38.0%; Score 272.5; DB 2; Length 225;
Best Local Similarity 59.8%; Pred. No. 2.9e-19;
Matches 55; Conservative 14; Mismatches 18; Indels 5; Gaps 2;

QY 16 KQARFGVRRPWGKFAAIRDPSRNGARLWLTGTETAEAAAYDRAAFNLRGHILAIN 75

DB 101 KGHYRGVRRPWGKFAAIRDPAKNGARVWLTGTETAEADALAYDRAAFMRGSRALLN 160

QY 76 FP---NEYYPMDYSLRPPYASSSSSSSGS 104

DB 161 FFLRVNSGEP--DPVRIKRSFSSNENGA 190

RESULT 7

T51988

ethylene responsive element binding factor 1 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000

C:Accession: T51988

R:Fujimoto, S.Y.; Ohta, M.; Usui, A.; Shinshi, H.; Ohme-Takagi, M.

Plant Cell 12, 393-404, 2000

A:Title: Arabidopsis ethylene responsive element binding factors act as transcriptional

A:Reference number: Z25893

A:Accession: T51988

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-266 <FUJ>

A:Cross-references: EMBL:AB008103; PIDN:BAA32418.1

C:Genetics:

A:Gene: ERF-1

Query Match 38.0%; Score 272.5; DB 2; Length 266;
Best Local Similarity 59.8%; Pred. No. 3.5e-19;
Matches 55; Conservative 14; Mismatches 18; Indels 5; Gaps 2;

QY 16 KQARFGVRRPWGKFAAIRDPSRNGARLWLTGTETAEAAAYDRAAFNLRGHILAIN 75

DB 142 KGHYRGVRRPWGKFAAIRDPAKNGARVWLTGTETAEADALAYDRAAFMRGSRALLN 201

QY 76 FP---NEYYPMDYSLRPPYASSSSSSSGS 104

DB 202 FFLRVNSGEP--DPVRIKRSFSSNENGA 231

RESULT 8

T07689

transcription factor Pti5 - tomato

C:Species: Lycopersicon esculentum (tomato)

C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 11-May-2000

C:Accession: T07689

R:Zhou, J.; Tang, X.; Martin, G.B.

EMBO J. 16, 3207-3218, 1997

A:Title: The Pto kinase conferring resistance to tomato bacterial speck disease interact

A:Reference number: Z16096; MUID:9757308; PMID:9214637

A:Accession: T07689

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-161 <ZHO>

A:Cross-references: EMBL:U89256; NID:92213782; PIDN:AAC49740.1; PID:92213783

C:Function:

A:Description: transcription factor

A>Note: binds the GCC box, present in the promoter region of genes encoding pathogenesis

C:Keywords: DNA binding; transcription factor

Query Match 37.9%; Score 272; DB 2; Length 161;
Best Local Similarity 53.8%; Pred. No. 2.2e-19;
Matches 56; Conservative 19; Mismatches 23; Indels 6; Gaps 2;

QY 19 RFRGVRPPWPKFAAETRDPSPNGARLWLTGTETAEAAAYDRAAFNLRGHILAINFPN 78

DB 58 KYRGVRRPWGKFAAETRDSARHGARVWLTGTETAEAAALAYDRAAFMRGAKALLNFP 117

QY 79 EYI---PRMDDYSLRPPYASSSSSGSTSTNVSRQNRQREVFE 119

DB 118 EIVNASVSDKLSL---CSNYYTTNNSSDLSNEVSSGNTNDVFE 158

RESULT 9

T00409

ethylene-responsive transcription factor homolog T13E15.15 - Arabidopsis thaliana

N:Alternate names: hypothetical protein At2g44840

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 16-Feb-2001

C:Accession: T00409; D84883

R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Masor

submitted to the EMBL Data Library, July 1997

A:Description: Arabidopsis thaliana chromosome II BAC T13E15 genomic sequence.

A:Reference number: Z14146

A:Accession: T00409

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-226 <ROU>

A:Cross-references: EMBL:AC002388; NID:93420042; PID:92344900

A:Experimental source: cultivar Columbia

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; N

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umavam, L.; Tallon, L.C.

euss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: D84883

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-226 <STO>

A:Cross-references: GB:AB002093; NID:92344900; PIDN:AAC31840.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g44840; T13E15.15

A:Map position: 2

Query Match 37.9%; Score 272; DB 2; Length 226;
Best Local Similarity 48.0%; Pred. No. 3.2e-19;
Matches 61; Conservative 18; Mismatches 30; Indels 18; Gaps 4;

QY 4 SNRSSNQSDDKQARFGRVRRPWGKFAAETRDPSPNGARLWLTGTETAEAAAYDRA 63

DB 76 ATKASGSHAPROKQMGVGRVRRPWGKFAAETRDPSPNGARVWLTGTETPDAAAYDRA 135

QY 64 AFNLRGHILAINFPN-----EYPRMDDYSLRPPYASSSSSGSTSTNVSRQNRQREVFE 118

DB 136 AFOLAGSKAKLNFPHLIGSCKYEP---VRIPRRSRPEPSVSDQLTS-----EQKRE-- 184

QY 119 EFEYLD 125

DB 185 --SHVDD 189

RESULT 10

T02590

DNA binding protein EREBP-2 - common tobacco

C:Species: Nicotiana tabacum (common tobacco)

C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 21-Jul-2000

C:Accession: T02590

R:Ohme-Takagi, M.; Shinshi, H.

Plant Cell 7, 173-182, 1995

A:Title: Ethylene-inducible DNA binding proteins that interact with an ethylene responsi

A;Reference number: Z14671; UID:95276459; PMID:7756828
A;Accession: T02590
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-233 <OHM>
A;Cross-references: EMBL:D38126; NID:G790362; PIDN:BAA07324.1; PID:g1208498
A;Experimental source: strain BY4; tissue-type leaf

Query Match 37.5%; Score 269.5; DB 2; Length 233;
Best Local Similarity 59.8%; Pred. No. 5.9e-19;
Matches 55; Conservative 12; Mismatches 20; Indels 5; Gaps 1;
QY 16 KOARFRGVRPPGKFAAEIRDSRNGARLWLTFTAEAAAYDRAAFNLRGHLAIN 75
DB 95 KGRHYRGVRQRPWGKFAAEIRDPKNGARVWLTFTAEAAAYDRAAFNLRGSKALLN 154

QY 76 FP-----NEYPRMDDYSLRPPYASSSSSSSSSS 102
DB 155 FPHRIGLNEPEPVRLTAKRRSPSPASSSSISSA 186

RESULT 11

T02434
DNA binding protein ERBBP-4 - common tobacco
C;Species: Nicotiana tabacum (common tobacco)
C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 31-Mar-2001
C;Accession: T02434
R;Ohme-Takagi, M.; Shinshi, H.
Plant Cell 7, 173-182, 1995
A;Title: Ethylene-inducible DNA binding proteins that interact with an ethylene response
A;Reference number: Z14671; UID:95276459; PMID:7756828
A;Accession: T02434
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-291 <OHM>
A;Cross-references: EMBL:D38125; NID:G790361; PIDN:BAA07323.1; PID:g1208497.
A;Experimental source: strain BY4; tissue-type leaf
C;Superfamily: Arabidopsis thaliana hypothetical protein T19P19.170

Query Match 36.9%; Score 265; DB 2; Length 291;
Best Local Similarity 46.0%; Pred. No. 2.1e-18;
Matches 58; Conservative 17; Mismatches 25; Indels 26; Gaps 3;
QY 16 KOARFRGVRPPGKFAAEIRDSRNGARLWLTFTAEAAAYDRAAFNLRGHLAIN 75
DB 141 EKHYRGVRQRPWGKFAAEIRDPNKGTRVWLTFTAEAAAYDRAAFNLRGSKRAIVN 200
QY 76 FPNFY--YPRMDDYSLRPPYASSSSSSSSSGSTSTNVSRQNVFEPEYLDKVLLELD 133
DB 201 FPLEVANFKQDNEILQ-----ANSGRKMRTE-----EEIVI 236

QY 134 SEERKR 139

DB 237 KKEVKR 242

RESULT 12

T52189
ethylene responsive element binding factor [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C;Accession: T52189
R;Usui, A.; Ohta, M.; Shinshi, H.; Ohme-Takagi, M.
submitted to the EMBL Data Library, April 1998
A;Description: Arabidopsis ethylene responsive element binding factor (ATERFs).
A;Reference number: Z25994
A;Accession: T52189
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-281 <USU>
A;Cross-references: EMBL:AB013301; PIDN:BAA31525.1
C;Genetics:
A;Gene: ERP6

Query Match 36.5%; Score 262; DB 2; Length 281;
Best Local Similarity 45.6%; Pred. No. 4e-18;
Matches 57; Conservative 18; Mismatches 24; Indels 26; Gaps 3;
QY 15 DKQARFRGVRPPGKFAAEIRDPSPNGARLWLTFTAEAAAYDRAAFNLRGHLAIL 74
DB 132 EKRHYRGVRPPGKFAAEIRDPTRRGTRVWLTFTAEAAAYDRAAFNLRGSKAIL 191
QY 75 NFP--NEYPRMDDYSLRPPYASSSSSSSSSGSTSTNVSRQNVFEPEYLDKVLLEL 131
DB 192 NFPLEVDKKNPRAED-----GRGLYKRKRDSSE-----BVKVVKV 228
QY 132 LQSEE 136
DB 229 LKKEE 233

RESULT 13

T52020
ethylene responsive element binding factor 5 [validated] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C;Accession: T52020
R;Fujimoto, S.Y.; Ohta, M.; Usui, A.; Shinshi, H.; Ohme-Takagi, M.
Plant Cell 12, 393-404, 2000
A;Title: Arabidopsis ethylene responsive element binding factors act as transcriptional
A;Reference number: Z25893
A;Accession: T52020
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-300 <FUJ>
A;Cross-references: EMBL:AB008107; PIDN:BAA32422.1
C;Genetics:
A;Note: ATERF-5
C;Function:
A;Description: acts as a transcriptional activator for GCC box-dependent transcription in

Query Match 36.5%; Score 262; DB 2; Length 300;
Best Local Similarity 65.4%; Pred. No. 4.3e-18;
Matches 51; Conservative 13; Mismatches 10; Indels 4; Gaps 2;
QY 12 SQDQKARFRGVRPPGKFAAEIRDPSPNGARLWLTFTAEAAAYDRAAFNLRGHL 71
DB 149 SEEEKK-HYRGVRQRPWGKFAAEIRDPNKGSRVWLTFTAEAAAYDRAAFNLRGSK 207
QY 72 AILNFPNE---YPRMDD 86
DB 208 AILNFPLEVGKWKPRADE 225

RESULT 14

E71444
probable ERBBP-4 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
A;Variety: columbia
C;Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 05-Dec-1998
C;Accession: E71444
R;Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirks
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weizengger, T.; Pohl, T.M.; Terry, N.; Giel
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
Nature 391, 485-488, 1998
A;Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech,
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Anst
C.; Chludzisz, N.
A;Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thali
A;Reference number: A71400; UID:98121113; PMID:9461215
A;Accession: E71444
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-603 <BEV>
A;Cross-references: GB:Z97343; NID:G22445073; PID:G227065; PID:G2245108
C;Genetics:

Search completed: December 17, 2003, 19:22:05
Job time : 28 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 18:23:03 ; Search time 18 Seconds
(without alignments)
363.151 Million cell updates/sec

Title: US-09-934-455-6
Perfect score: 718
Sequence: 1 MESSNRSSNQSDDKQARF.....PEYLDKVLBLEDSERKR 139

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 segs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	293	40.8	243	1	ERF2_ARATH
2	272.5	38.0	268	1	ERF1_ARATH
3	272	37.9	161	1	PT15_LYCES
4	262	36.5	300	1	ERF5_ARATH
5	244.5	34.1	225	1	ERF3_ARATH
6	240.5	33.5	248	1	AP23_ARATH
7	239.5	33.4	222	1	ERF4_ARATH
8	214.5	29.9	248	1	PT16_LYCES
9	161.5	22.5	164	1	P202_LUPPO
10	116.5	16.2	344	1	RAV1_ARATH
11	116.5	16.2	432	1	AP2_ARATH
12	91	12.7	179	1	PGSG_RAT
13	84.5	11.8	1677	1	VIT_ACITR
14	80	11.1	277	1	YMO5_CAEEL
15	78.5	10.9	1244	1	SLA1_YEAST
16	77	10.7	717	1	ARA_DROME
17	75.5	10.5	1337	1	YDM5_SCHPO
18	75	10.4	1259	1	NAH2_SCHPO
19	75	10.4	1532	1	IGA_NERGO
20	74.5	10.4	213	1	VNEN_PAVBO
21	74.5	10.4	955	1	T150_HUMAN
22	74	10.3	708	1	SSRP_MOUSE
23	74	10.3	1823	1	VIT_ICHUN
24	73.5	10.2	506	1	VE2_HPV47
25	73	10.2	843	1	CYPI_BRUMA
26	72.5	10.1	797	1	PAT1_YEAST
27	72.5	10.1	809	1	ENPL_HORVU
28	72.5	10.1	817	1	DLG3_HUMAN
29	72.5	10.1	849	1	DLG3_MOUSE
30	72.5	10.1	849	1	DLG3_RAT
31	72.5	10.1	2004	1	CHDB_HUMAN
32	72	10.0	658	1	CPT2_RAT
33	72	10.0	985	1	4ET_HUMAN

34	71.5	10.0	1912	1	VIT1_CHICK	P87498 gallus gall
35	71	9.9	559	1	FRB7_WHEAT	Q43207 triticum ae
36	70.5	9.8	372	1	YP99_CAEEL	Q09477 caenorhabdi
37	70	9.7	526	1	CLOS_CLOHI	P09870 clostridium
38	70	9.7	983	1	Y144_HUMAN	Q14157 homo sapien
39	69	9.6	1713	1	LMA3_HUMAN	Q16787 homo sapien
40	69	9.6	1901	1	YCF1_TOBAC	P12222 nicotiana t
41	68.5	9.5	486	1	KIC0_XENLA	P02537 xenopus lae
42	68	9.5	619	1	T7L2_HUMAN	Q09q00 homo sapien
43	68	9.5	1162	1	ITAD_HUMAN	Q13349 homo sapien
44	68	9.5	1687	1	VIT2_FUNHE	Q98893 fundulus he
45	67.5	9.4	537	1	YDU2_SCHPO	O13863 schizosacch

ALIGNMENTS

RESULT 1

ID	ERF2_ARATH	STANDARD;	PRT;	243 AA.
AC	O80338;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Ethylene responsive element binding factor 2 (ATERF2).			
GN	ERF2 OR ERF-2 OR AT5G47220 OR MQL5.7.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	euroside II; Brassicales; Brassicaceae; Arabidopsis.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20181733; PubMed=10715325;			
RA	Fujimoto S.Y., Ohta M., Usui A., Shinshi H., Ohme-Takagi M.;			
RT	"Arabidopsis ethylene responsive element binding factors act as			
RT	transcriptional activators or repressors of GCC box mediated gene			
RT	expression.";			
RL	Plant Cell 12:393-404(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Columbia;			
RX	MEDLINE=20181125; PubMed=10718197;			
RA	Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,			
RA	Tabata S.;			
RT	"Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence			
RT	features of the regions of 3,076,755 bp covered by sixty P1 and TAC			
RT	clones.";			
RL	DNA Res. 7:31-63(2000).			
CC	-!- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS TO THE GCC-BOX			
CC	PATHOGENESIS-RELATED PROMOTER ELEMENT.			
CC	-!- SUBCELLULAR LOCATION: Nuclear (Probable).			
CC	-!- SIMILARITY: Contains 1 AP2/ERF domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AB008104; BAA32419.1; --			
DR	EMBL; AB018117; BAA97155.1; --			
DR	PIR; T51989; T51989.			
DR	HSSP; O80337; 2GCC.			
DR	TRANSFAC; T04634; --			
DR	InterPro; IPR001471; TF_ERF.			
DR	Pfam; PF00847; AP2-domain; 1.			
DR	ProDom; PD001423; TF_ERF; 1.			
DR	SMART; SM00380; AP2; 1.			
DR	Transcription regulation; DNA-binding; Nuclear protein.			
KW	DOMAIN 23 26			
FT	POLY-GLY.			


```
DR EMBL; AB008103; BAA32418.1; ALT INIT.
DR EMBL; Z97343; CAB45963.1; ALT FRAME.
DR EMBL; AL161546; CAB78753.1; ALT FRAME.
DR EMBL; AY058174; AAL25588.1; -.
DR EMBL; AY062533; AAL32611.1; -.
DR PDB; 1GCC; 23-MAR-99.
DR PDB; 2GCC; 23-MAR-99.
DR PDB; 3GCC; 23-MAR-99.
DR TRANSFAC; T04633; -.
DR InterPro; IPR001471; TF_ERF.
DR Pfam; PF00847; AP2-domain; 1.
DR PRINTS; PRO0367; ETHRSPLEKMT.
DR ProDom; PD001423; TF_ERF; 1.
DR SMART; SM00380; AP2; 1.
DR Transcription regulation; DNA-binding; Activator; Nuclear protein.
DR Transcription regulation; DNA-binding; Activator; Nuclear protein.
FT DNA_BIND 56 121
SQ SEQUENCE 161 AA; 18051 MW; 67F4D2D4C0B37FB0 CRC64;

Query Match 37.9%; Score 272; DB 1; Length 161;
Best Local Similarity 53.8%; Pred. No. 2.9e-19;
Matches 56; Conservative 19; Mismatches 23; Indels 6; Gaps 2;

QY 19 RFRGVRPRPWGKFAAEIRDPNRNGARLWLTETAEAAAYDRAAFNLRGHLAIIIN 78
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
58 KYRGVRRPRPWGKFAAEIRDPNRNGARLWLTETAEAAAYDRAAFNLRGHLAIIIN 117
QY 79 EYV---PRMDDYSLRPPYASSSSSSSSSGSTSTNVSRQNRQVEE 119
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
118 EIVNASVSVDKLSL---CSNYYTTNNNSDSSLNVSSGCTNDVFE 158

RESULT 4
ERFS_ARATH STANDARD; PRT; 300 AA.
AC 080341;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ethylene responsive element binding factor 5 (ATERF5).
GN ERFS OR ERP-5 OR AT5G47230 OR MQL5_9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20181733; PubMed=10715325;
RA Fujimoto S.Y., Ohta M., Utsui A., Shinsui H., Ohme-Takagi M.;
RT "Arabidopsis ethylene responsive element binding factors act as
RT transcriptional activators or repressors of GCC box mediated gene
RT expression."
RL Plant Cell 12:393-404 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
RT SSP consortium (Salk/Stanford/PGECC).";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS TO THE GCC-BOX
CC -!- PATHOGENESIS-RELATED PROMOTER ELEMENT.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: Contains 1 AP2/ERF domain.
CC -!- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -!- EMBL; AB008107; BAA32422.1; -.
CC -!- EMBL; AF385709; AAK60301.1; -.
CC -!- PIR; T52020; T52020.
CC -!- HSSP; O80337; 2GCC.
CC -!- TRANSFAC; T04637; -.
CC -!- InterPro; IPR001471; TF_ERF.
CC -!- Pfam; PF00847; AP2-domain; 1.
CC -!- ProDom; PD001423; TF_ERF; 1.
CC -!- SMART; SM00380; AP2; 1.
CC -!- Transcription regulation; DNA-binding; Nuclear protein.
KW
```

```
FT DOMAIN 51 55 POLY-SER.
FT DNA BIND 153 218 AP2/ERF.
SQ SEQUENCE 300 AA; 33810 MW; 1189D46A28F7251 CRC64;

Query Match 36.5%; Score 262; DB 1; Length 300;
Best Local Similarity 65.4%; Pred. No. 5.7e-18;
Matches 51; Conservative 13; Mismatches 10; Indels 4; Gaps 2;

QY 12 SODKQARFRGVRPPWKGFAAEIRDPNRGALRWLGTFTAEAAAYDRAAFNLRGHL 71
DB 149 SEEEK-HYRGVRPPWKGFAAEIRDPNRGALRWLGTFTAEAAAYDRAAFNLRGHL 207

QY 72 AILNFPNE---YYPRMD 86
DB 208 AILNFPLEVGKWKPRADE 225

RESULT 5
ERF3 ARATH STANDARD; PRT; 225 AA.
AC O80339;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ethylene responsive element binding factor 3 (ATERF3).
GN ERF3 OR ERF-3 OR AT1G50640 OR F17J6.16 OR F11F12.4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20181733; PubMed=10715325;
RA Fujimoto S.Y., Ohta M., Utsi A., Shinshi H., Ohme-Takagi M.;
RT "Arabidopsis ethylene responsive element binding factors act as
RT transcriptional activators or repressors of GCC box mediated gene
RL expression."
RL Plant Cell 12:393-404(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altai H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Bucher E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.P., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Millescher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana."
RL Nature 408:816-820(2000).
CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS TO THE GCC-BOX
CC -1- PATHOGENESIS-RELATED PROMOTER ELEMENT.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: Contains 1 AP2/ERF domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/)
```

```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB008105; BAA32420.1; -
DR EMBL; AC079279; AAG51201.1; -
DR EMBL; AC012561; AAF87871.1; -
DR PIR; T52011; T52011.
DR HSSP; O80337; 2GCC.
DR TRANSFAC; T04635; -
DR InterPro; IPR001471; TF_ERF.
DR Pfam; PF00847; AP2-domain; 1.
DR PRINTS; PRO0367; ETHRSPPELEMT.
DR ProDom; PD001423; TF_ERF; 1.
DR SMART; SM00380; AP2; 1.
KW Transcription regulation; DNA-binding; Nuclear protein.
FT DNA BIND 25 89 AP2/ERF.
FT DOMAIN 90 93 POLY-PRO.
FT DOMAIN 175 178 POLY-SER.
FT DOMAIN 181 185 POLY-ASP.
SQ SEQUENCE 225 AA; 25216 MW; C709BEAFB306776B CRC64;

Query Match 34.1%; Score 244.5; DB 1; Length 225;
Best Local Similarity 47.1%; Pred. No. 1.9e-16;
Matches 57; Conservative 13; Mismatches 24; Indels 27; Gaps 3;

QY 16 KQARFRGVRPPWKGFAAEIRDPNRGALRWLGTFTAEAAAYDRAAFNLRGHLAIN 75
DB 24 KEIRFGVRPPWKGFAAEIRDPWKK-ARVWLGTFTDSAEAAAYDRAAFNLRGPKAKTN 82

QY 76 FP-----NEYPRMD-----DYSLRPPYASSSSSSSGSGSTNV 109
DB 83 FPIDSSPPPPNLRFRNQNRQNVDPFMDHRLFTDHOQQQFPVNRPTSSSSSTVESF 142

QY 110 S 110
DB 143 S 143

RESULT 6
AP23 ARATH STANDARD; PRT; 248 AA.
AC P42736; O23105;
DT 01-NOV-1995 (Rel. 32, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE AP2 domain transcription factor RAP2.3 (Related to AP2 protein 3)
DE (Cadmium-induced protein AS30).
GN AT3G16770 OR MGL6.1 OR MGL6.24.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX Choi S., Baek E., Lee S.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97338147; PubMed=9192694;
RA Okumuro J.K., Caster B., Villarroel R., van Montagu M., Jofuku K.D.;
RT "The AP2 domain of APTALA2 defines a large new family of DNA binding
RT proteins in Arabidopsis."
RL Proc. Natl. Acad. Sci. U.S.A. 94:7076-7081(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=cv. Columbia;
RX MEDLINE=20277480; PubMed=10819329;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones."
RL DNA Res. 7:131-135(2000).
```



```

RN RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "RIKEN Arabidopsis full length cDNA clones (RAPFs) sequenced by the
RL SSP consortium (Salk/Stanford/PGEC).";
RN [5] Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
RN RP SEQUENCE FROM N.A.
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -I- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -I- INDUCTION: By cadmium.
CC -I- SIMILARITY: Contains 1 AP2/ERF domain.
CC -I- CAUTION: Ref.1 sequence is most probably a cloning artifact and
CC differs from that shown due to an insertion into the sequence.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z37504; CAAG5734.1; ALT_SEQ.
DR EMBL; A7003096; AAC49769.1; -.
DR EMBL; AB022217; BAB02769.1; -.
DR EMBL; AY059917; AAL24399.1; -.
DR EMBL; AY035100; AAK59605.1; -.
DR EMBL; AV142562; AAN13131.1; -.
DR EMBL; AV087488; AAM65031.1; -.
DR HSP; O80337; 2GCC.
DR TRANSFAC; T02658; -.
DR InterPro; IPR001471; TF_ERF.
DR Pfam; PF00847; AP2-domain; 1.
DR PRINTS; PR00367; ETHRSPELEMT.
DR ProDom; PD001423; TF_ERF; 1.
DR SMART; SM00380; AP2; 1.
DR Transcription regulation; DNA-binding; Nuclear protein; Cadmium.
FT DNA_BIND 76 141
FT DOMAIN 140 143 POLY-PRO.
SQ SEQUENCE 248 AA; 27758 MW; 317D50CC514782F1 CRC64;

Query Match 33.5%; Score 240.5; DB 1; Length 248;
Best Local Similarity 51.0%; Pred. No. 5.2e-16;
Matches .50; Conservative 17; Mismatches 28; Indels 3; Gaps 2;

QY 20 FRCVRRPWGKFAAEIRDPSSRCARLWLGTFETAEAAAYDPAAFNLGHLAINFPNE 79
DB 79 YGIRKRPWGWAAEIRDP-RKGVRLWLGTFNTAEAAAYDVAAKQIRGDKAKLNFDDL 137
QY 80 YPRMDYSLRPYASSSSSGSTSTNSVRQNGREV 117
DB 138 HHPPPNYT--PPSPSRSTDQPPAKKVCVVSQSESEL 173

RESULT 7
ERF4_ARATH
ID _ERF4_ARATH STANDARD; PRT; 222 AA.
AC O80337;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ethylene responsive element binding factor 4 (AtERF4).
GN ERF4 OR ERF-4 OR AT3G5210 OR K7L4.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;

```

```

RN RP SEQUENCE FROM N.A.
RX MEDLINE=20181733; PubMed=10715325;
RA Fujimoto S.Y., Ohta M., Usui A., Shinshi H., Ohme-Takagi M.;
RT "Arabidopsis ethylene responsive element binding factors act as
RT transcriptional activators or repressors of GCC box mediated gene
RT expression.";
RL Plant Cell 12:393-404(2000).
RN [2]
RN RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016720; PubMed=11130713;
RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unselid M.,
RA Partmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
RA Delseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA De Simone V., Choisine N., Artiguenave F., Robert C., Bottier P.,
RA Wincker P., Cattolico L., Weissenbach J., Saurin W., Queirer F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wurmbach E., Drzonek H., Erile H., Jordan N., Bangert S.,
RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,
RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordsiek G.,
RA Reichelt J., Scharfe M., Schoen O., Barges M., Terol J., Climent J.,
RA Navarro P., Collado C., Perez-Perez A., Pottenwaelder B., Duchemin D.,
RA Cooke R., Laudie M., Berger-Liauro C., Furnelle B., Masuy D.,
RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA Monfort A., Agirion A., Flores M., Liguori R., Vitale D.,
RA Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Waits A., Utterback T., Fujii C.Y., Shea T.P.,
RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA Pai G., Militscher J., Sellers P., Gill J.E., Feldblyum T.V.,
RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT thaliana".
RL Nature 408:820-822(2000).
CC -I- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS TO THE GCC-BOX
CC PATHOGENESIS-RELATED PROMOTER ELEMENT.
CC -I- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -I- SIMILARITY: Contains 1 AP2/ERF domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB008106; BAA32421.1; -.
DR EMBL; AP000413; BAB02150.1; -.
DR PIR; T52019; T52019.
DR HSP; O80337; 2GCC.
DR TRANSFAC; T04636; -.
DR InterPro; IPR001471; TF_ERF.
DR Pfam; PF00847; AP2-domain; 1.
DR PRINTS; PR00367; ETHRSPELEMT.
DR ProDom; PD001423; TF_ERF; 1.
DR SMART; SM00380; AP2; 1.
DR Transcription regulation; DNA-binding; Nuclear protein.
FT DNA_BIND 22 86
FT DOMAIN 132 137 POLY-GLY.
SQ SEQUENCE 222 AA; 23741 MW; CEF16ABA5ACFB073 CRC64;

Query Match 33.4%; Score 239.5; DB 1; Length 222;
Best Local Similarity 44.3%; Pred. No. 5.7e-16;
Matches 54; Conservative 20; Mismatches 29; Indels 19; Gaps 4;

```

```
Qy 3 SSNRSSNNQDDKQARFGRVRRPWWGKFAAEIRDPNRGRLWLGTFTAEAAAYDR 62
Db 12 TTQTHNNA-----KEIRYRGVRKRPWGRYAAEIRDPGKK-TRVWLGTFTAEAAAYDT 66

Qy 63 AAFNLRGHLAALNPNPNYPRMDYSLRPPYASSSSSSSS-----GSTSTNVS 110
Db 67 AARDFRGAKAKTNFT--FUELSQKQVPTGFARSPQSSTLDCASPTLVVPVSATAGNP 124

Qy 111 RQ 112
Db 125 PQ 126

RESULT 8
PT16_LYCES
ID PT16_LYCES STANDARD; PRT; 248 AA.
AC 004682;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Pathogenesis-related genes transcriptional activator PT16.
GN PT16.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiids; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97357308; PubMed=9214637;
RA Zhou J., Tang X., Martin G.B.;
RT "The Pto kinase conferring resistance to tomato bacterial speck
RT disease interacts with proteins that bind a cis-element of
RT pathogenesis-related genes.";
RL EMO J. 16:3207-3218 (1997).
CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS TO THE GCC-BOX
CC -1- PATHOGENESIS-RELATED PROMOTER ELEMENT. ACTIVATES PLANTS DEFENSE
CC GENES.
CC -1- SUBUNIT: INTERACTS WITH THE PTO KINASE.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: Contains 1 AP2/ERF domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U89257; AAC49741.1; -
CC PIR; T07728; T07728.
CC HSP; O80337; 2GCC.
CC TRANSFAC; T04717; -.
CC InterPro; IPR001471; TF ERF.
CC Pfam; PF00847; AP2-domain; 1.
CC ProDom; PD001423; TF ERF; 1.
CC SMART; SM00390; AP2_1.
CC Transcription regulation; DNA-binding; Activator; Nuclear protein.
CC DNA_BIND 95 159 AP2/ERF.
CC SEQUENCE 248 AA; 27913 MW; 115BD4DDEB1B992 CRC64;

Query Match 29.9%; Score 214.5; DB 1; Length 248;
Best Local Similarity 53.9%; Pred. No. 1.6e-13;
Matches 41; Conservative 15; Mismatches 19; Indels 1; Gaps 1;

Qy 2 SSNRSSNNQDDKQARFGRVRRPWWGKFAAEIRDPNRGRLWLGTFTAEAAAYD 61
Db 80 DRKRSVSPSDVTRKKFGVRQRPWGRWAAEIRDPTR-GKRVWLGTFTPEAAVVD 138

Qy 62 RAAPNLRGHLAALNPNP 77
Db 139 KAAVKLGPDAVINFP 154
```

RESULT 9

```
PZ02_LUPPO
ID PZ02_LUPPO STANDARD; PRT; 164 AA.
AC P16146;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PZL202 protein.
GN PZL202.
OS Lupinus polyphyllus (large-leaved lupine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Genistaceae; Lupinus.
OX NCBI_TaxID=3874;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91355854; PubMed=2103434;
RA Perrey R., Warskulat U., Wink M.;
RT "Molecular cloning of a lupin-specific gene from a cDNA library of
RT suspension-cultured cells of Lupinus polyphyllus.";
RL Plant Mol. Biol. 15:175-176 (1990).
CC -1- FUNCTION: ESSENTIAL FOR ALL LUPIN CELLS INDEPENDENT OF THE
CC RESPECTIVE TISSUE.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: Contains 1 AP2/ERF domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X51767; CAA36069.1; -
CC PIR; S11881; S11881.
CC HSP; O80337; 2GCC.
CC InterPro; IPR001471; TF ERF.
CC Pfam; PF00847; AP2-domain; 1.
CC PRINTS; PR00367; ETRSPLEMT.
CC ProDom; PD001423; TF ERF; 1.
CC SMART; SM00380; AP2_1.
CC Transcription regulation; DNA-binding; Nuclear protein.
CC DNA_BIND 4 69 AP2/ERF.
CC SEQUENCE 164 AA; 18606 MW; 87BF64C0122ECB00 CRC64;

Query Match 22.5%; Score 161.5; DB 1; Length 164;
Best Local Similarity 48.8%; Pred. No. 1.2e-08;
Matches 41; Conservative 9; Mismatches 21; Indels 13; Gaps 3;

Qy 17 QARFGRVRRPWWGKFAAEIRDPNRGRLWLGTFTAEAAAYDRPAAFNLRGHLALNF 76
Db 5 QQRYGRGVRQHWGWSVSEIRH-SILKTRIQGTFTSAEDARAYDEARLMCGTRATNF 63

Qy 77 PNEYPRMDYSLRPPYASSSSSS 100
Db 64 P-----YN---PNASQSSSS 75

RESULT 10
RAV1_ARATH
ID RAV1_ARATH STANDARD; PRT; 344 AA.
AC Q9ZMM9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA-binding protein RAV1.
GN RAV1 OR ATG13260 OR T6J4.2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
```

OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 RN NCBI_TaxID=3702;
 RP SEQUENCE FROM N.A.
 RC SEQUENCE Seedling; PubMed=9862967;
 RX MEDLINE=99081843; PubMed=9862967;
 RA Kagaya Y., Ohmura K., Hattori T.;
 RT "RAVI, a novel DNA-binding protein, binds to bipartite recognition
 RT sequence through two distinct DNA-binding domains uniquely found in
 RT higher plants.";
 RL Nucleic Acids Res. 27:470-478 (1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cnv. Columbia;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altai H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Egtu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
 RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzer S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Utterback T., Van Aken S., Vaynsberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 408:816-820 (2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cnv. Columbia;
 RX Shinzaki K., Davis R.W., Ecker J.R., Theologis A.;
 RA "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
 RT SSP consortium (Salk/Stanford/PGECC).";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Binds specifically to bipartite recognition sequences
 CC composed of two unrelated motifs, 5'-CAACA-3' and 5'-CACCTG-3'.
 CC -!- SUBUNIT: Monomer.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- TISSUE SPECIFICITY: Expressed in all tissues examined. Roots,
 CC rosette leaves, cauline leaves, inflorescence stems, flowers and
 CC siliques. Highest expression in roots and rosette leaves. Very low
 CC expression in flowers.
 CC -!- DOMAIN: Contains two distinct DNA-binding domains. One is located
 CC in the N-terminal region and binds to the 5'-CAACA-3' motif. The
 CC second is located in the C-terminal region and binds to the
 CC 5'-CACCTG-3' motif.
 CC -!- SIMILARITY: Contains 1 AP2/ERF domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB013886; BAA34250.1; -;
 DR EMBL; AC011810; AAG09554.1; -;
 DR EMBL; AY063855; AAL36211.1; -;
 DR EMBL; AY091291; AAM14230.1; -;
 DR F1; T51329; T51329;
 DR HSP; O80337; 2GCC.
 DR InterPro; IPR003340; TF.B3.
 DR InterPro; IPR001471; TF.BRF.
 DR Pfam; PF00847; AP2-domain; 1.
 DR Pfam; PF02362; B3; 1.

DR PRINTS; PR00367; ETRSRPELEMT.
 DR ProDom; PD001423; TF.BRF; 1.
 KW DNA-binding; Nuclear protein.
 FT DNA_BIND 59 121 AP2/ERF.
 FT DOMAIN 145 338 B3.
 SQ SEQUENCE 344 AA; 38597 MW; 7349B640B3505823 CRC64;
 Query Match 16.2%; Score 116.5; DB 1; Length 344;
 Best Local Similarity 27.9%; Pred. No. 0.00061;
 Matches 36; Conservative 26; Mismatches 42; Indels 25; Gaps 5;
 QY 11 OSODDKQARFRGVRPPWKGFAAEIRDPNRGALWLTGTTAEAAAYDAAFNLRGH 70
 DB 53 ESRKLPSSKYGVVQPNRGAQIYERHQ---RVMLGTENEDEAAAYDAVHRFR 109
 QY 71 LAIINFNPYPRMDYSLRPPYASSSSSSSGSTSTNVSRQNRQEFVEFYLDKVL 130
 DB 110 DAVTNFKD---VKMDEDEV--DFLNSHKS-----EIVDM--LRKHTYNE 147
 QY 131 LLDSEERKR 139
 DB 148 ELEQSKRRR 156
 RESULT 11
 ID AP2_ARATH STANDARD; PRT; 432 AA.
 AC P47327;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Floral homeotic protein APETALA2.
 GN AP2 OR AT4G36920 OR AP22.49 OR C7A10.440.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A., AND MUTANTS AP2-1 AND AP2-5.
 RC STRAIN=cnv. Columbia; TISSUE=Flower;
 RX MEDLINE=95003709; PubMed=7919989;
 RA Jofuku K.D., den Boer B.G.W., van Montagu M., Okamura J.K.;
 RT "Control of Arabidopsis flower and seed development by the homeotic
 RT gene APETALA2.";
 RL Plant Cell 6:1211-1225 (1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cnv. Columbia;
 RX MEDLINE=20083488; PubMed=10617198;
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terry N.,
 RA Harris B., Ansgore W., Brandt P., Grivell L.A., Rieger M.,
 RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
 RA Kreis M., Delsen T., Puigdomenech P., Watson M., Schmidheini T.,
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.A., McCullagh B., Bilham L., Robben J.,
 RA Van der Schueren J., Grynoprez B., Chuang Y.-J., Vandenbussche F.,
 RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
 RA Weizenecker T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirks W.,
 RA Moolman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
 RA Bernier S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
 RA De Keyser A., Buysschaert C., Giesen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIlroy K., Mayes R.,
 RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
 RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fartmann B., Grandrath K., Dauner D., Herzl A.,
 RA Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,
 RA Massenot O., Quigley F., Ciabaud G., Muendlein A., Felber R.,

RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Chedford F., Cooke R., Berger C., Monfort A., Casacuberta E.,
 RA Gibbons T., Weber N., Vandenbol M., Barques M., Terol J., Torres A.,
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Frances P., Bielke C.,
 RA Frishman D., Haase D., Lencke K., Mewes H.-W., Stocker S.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
 RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
 RA Latreille P., Courtnay L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Mardis E., Dante M., Peco T., Hillier L.,
 RA Nelson J., Spiech J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshua C.,
 RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
 RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
 RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
 RA Chen E., Marra M., Martienssen R., McCombie W.R.,
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 RT thaliana";
 RL Nature 402:769-777(1999).
 RN [3]
 RP FUNCTION.
 RP PubMed=1675158;
 RA Drews G.N., Bowman J.L., Meyerowitz E.M.;
 RT "Negative regulation of the Arabidopsis homeotic gene AGAMOUS by the
 RT APETALA2 product.";
 RL Cell 65:991-1002(1991).
 CC -!- FUNCTION: Probable transcriptional activator that promotes early
 CC floral meristem identity. Is required subsequently for the
 CC transition of an inflorescence meristem into a floral meristem.
 CC Plays a central role in the specification of floral identity,
 CC particularly for the normal development of sepals and petals in the
 CC wild-type flower. Acts as a class A class cadasteral protein by repressing
 CC the C class floral homeotic gene AGAMOUS in association with repressing
 CC other repressors like LEUNIG and SEUSS. It is also required
 CC during seed development.
 CC -!- SUBUNIT: May form homodimer.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- TISSUE SPECIFICITY: Sepals, petals, stamens, carpels, developing
 CC ovules, inflorescence stem, leaf, and stem.
 CC -!- DEVELOPMENTAL STAGE: It is detectable at low levels throughout the
 CC shoot apex, and at enhanced levels in the inflorescence meristem,
 CC young floral buds, and throughout the early stages of flower
 CC development and organogenesis. During floral organ differentiation
 CC it becomes spatially restricted to specific organ, tissue, and
 CC cell types within the flower.
 CC -!- INDUCTION: Negatively regulated by the C class floral
 CC homeotic protein AGAMOUS in stamens and carpels.
 CC -!- MISCELLANEOUS: Mutations in the APETALA2 gene result in the
 CC ectopic expression of AGAMOUS, leading to the replacement of
 CC sepals by carpels and stamens and of petals by stamens.
 CC -!- SIMILARITY: Contains 2 AP2/ERF domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U12546; AAC13770.1; -;
 DR EMBL; Z99707; CAB16765.1; -;
 DR EMBL; AL161590; CAB80358.1; -;
 DR PIR; A85436; A85436.
 DR TRANSFAC; T01774; -;
 DR InterPro; IPK001471; TF_ERF.
 DR Pfam; PF00847; AP2-domain; 2.
 DR PRINTS; PR00367; ETHRSPLEMT.
 DR ProDom; PD001423; TF_ERF; 1.

DR SMART; SM00380; AP2; 2.
 KW Flowering; Transcription regulation; Activator; Developmental protein;
 KW Nuclear protein; DNA-binding; Repeat.
 FT DOMAIN 14 50 ASP/GLU/SER-RICH (ACIDIC; POTENTIAL
 FT INVOLVEMENT WITH TRANSCRIPTION).
 FT DOMAIN 119 128 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DNA_BIND 128 192 AP2/ERF 1.
 FT DNA_BIND 220 286 AP2/ERF 2.
 FT MUTAGEN 159 159 G->E: TEMPERATURE SENSITIVE; AP2-5.
 FT MUTAGEN 251 251 G->S: TEMPERATURE SENSITIVE; AP2-1.
 FT MUTAGEN 420 420 Q->E: TEMPERATURE SENSITIVE; AP2-5.
 SQ SSSEQUENCE 432 AA; 47833 MW; 376A355291EFFB42 CRC64;
 Query Match 16.2%; Score 116.5; DB 1; Length 432;
 Best Local Similarity 37.5%; Pred. No. 0.0008;
 Matches 33; Conservative 16; Mismatches 24; Indels 15; Gaps 4;
 OY 1 MESSNRSSNNQSDQKQAFRGVR-RRPMGKFAEIRDSRNGARLWGLTFETAEAAARA 59
 Db 118 LKSRGRGPRSR-----SQVRGTVFYRTGRWESHIWD---CGQVYLGFGDTAAHAARA 169
 OY 60 YDRAAFNLRLHLALNFPNEYPRMDDY 87
 Db 170 YDRAAIKFRGVADINF-----NIDY 191
 RESULT 12
 PGSG_RAT
 ID PGSG_RAT STANDARD; PRT; 179 AA.
 AC P04917;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Secretory granule proteoglycan core protein precursor (Chondroitin
 DE sulfate proteoglycan core protein) (Proteoglycan 10K core protein).
 DE (PG19 core protein) (Cytolytic granule proteoglycan core protein).
 GN PRG1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=101116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88213412; PubMed=3366780;
 RA Avraham S., Stevens R.L., Gartner M.C., Austen K.F., Lalley P.A.,
 RA Weis J.H.;
 RT "Isolation of a cDNA that encodes the peptide core of the secretory
 RT granule proteoglycan of rat basophilic leukemia-1 cells and
 RT assessment of its homology to the human analogue.";
 RL J. Biol. Chem. 263:7292-7296(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Yolk sac carcinomas;
 RX MEDLINE=86304425; PubMed=2427521;
 RA Bourdon M.A., Shiga M., Ruoslahti E.;
 RT "Identification from cDNA of the precursor form of a chondroitin
 RT sulfate proteoglycan core protein.";
 RL J. Biol. Chem. 261:12534-12537(1986).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90279692; PubMed=2352541;
 RA Giorda R., Chambers W.H., Dahl C.A., Trucco M.;
 RT "Isolation and characterization of a cDNA that encodes the core
 RT protein of the cytolitic granule proteoglycan in rat natural killer
 RT cells.";
 RL Nat. Immun. Cell Growth Regul. 9:91-102(1990).
 RN [4]
 RP SEQUENCE OF 75-179 FROM N.A.
 RX MEDLINE=85140287; PubMed=3919394;
 RA Bourdon M.A., Oldberg A., Pierschbacher M.D., Ruoslahti E.;
 RT "Molecular cloning and sequence analysis of a chondroitin sulfate
 RT proteoglycan cDNA.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:1321-1325(1985).

```

CC -!- SIMILARITY: TO CORRESPONDING PROTEINS IN HUMAN AND MOUSE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: K02934; AAA42171.1; -.
CC EMBL: J03224; AAA41837.1; -.
CC PIR: A25644; GZRT0.
CC Pfam: PF04360; Serglycin; 1.
CC Glycoprotein; Proteoglycan; Repeat; Signal.
CC SIGNAL 1 26 POTENTIAL.
CC PROPEP 27 75 ACTIVATION PEPTIDE.
CC CHAIN 76 179 SECRETORY GRANULE PROTEOGLYCAN CORE
CC PROTEIN.
CC DOMAIN 90 137 24 X 2 AA TANDEM REPEATS OF S-G.
CC DISULFID 39 48 POTENTIAL.
CC SEQUENCE 179 AA; 18577 MW; D2E2A8E7D3AAA0D6 CRC64;
Query Match 12.7%; Score 91; DB 1; Length 179;
Best Local Similarity 32.3%; Pred. No. 0.077;
Matches 32; Conservative 15; Mismatches 40; Indels 12; Gaps 3;
CC QY 8 SNNQSDDKQARFGRVRRPWGKFAAEIRDPSPNGARLWLGTFETAEAAAYDRAAFNL 67
CC DB 22 SSVQGPARRARYQWRCKPDGIFANCIEE---KGPRFDL-----IABES-----NVGPPM 69
CC QY 68 RGHILALNFPNEYPRMDYSLRPPYASSSSSSSGSGS 106
CC DB 70 TDPVLMRGFNDFPISDDYSGSGSGSGSGSGS 108
RESULT 13
ID VIT_ACITR STANDARD; PRT; 1677 AA.
AC Q90243;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Vitellogenin precursor (VTG) [Contains: lipovitellin I (LVI);
DE Phosvitin (PV); Lipovitellin II (LVI)] (Fragment).
OS Acipenser transmontanus (White sturgeon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Chondrostei; Acipenseriformes; Acipenseridae;
OC Acipenser.
OX NCBI_TaxID=7904;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UC Davis Broodstock; TISSUE=Liver;
RX MEDLINE=95333192; PubMed=7608984;
RA Bidwell C.A., Carlson D.M.;
RT "Characterization of vitellogenin from white sturgeon, Acipenser
RT transmontanus."
RL J. Mol. Evol. 41:104-112(1995).
CC -!- FUNCTION: PRECURSOR OF THE MAJOR EGG-YOLK PROTEINS THAT ARE
CC SOURCES OF NUTRIENTS DURING EARLY DEVELOPMENT OF OVIPAROUS
CC ORGANISMS.
CC -!- TISSUE SPECIFICITY: FOUND IN LIVER, TESTIS AND UNDIFFERENTIATED
CC GONADS OF ESTROGEN-TREATED FISH. NOT DETECTED IN THE BRAIN AND
CC SPLEEN.
CC -!- INDUCTION: By steroids (estrogen).
CC -!- PTM: PHOSVITIN, AN EGG YOLK STORAGE PROTEIN, IS ONE OF THE MOST
CC HIGHLY PHOSPHORYLATED (10%) PROTEINS IN NATURE (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U0455; AAA87392.1; -.
CC PIR: T43021; T43021.
CC InterPro: IPR001747; Lipid_transprt_N.
CC InterPro: IPR001846; VWF_D.
CC Pfam: PF01347; Vitellogenin_N; 1.
CC Pfam: PF00094; vwd; 1.
CC SMART; SM00638; LPD_N; 1.
CC SMART; SM00216; VWD_1.
CC Glycoprotein; Phosphorylation; Storage protein; Signal.
CC NON_TER 1 1
CC SIGNAL <1 8 POTENTIAL.
CC CHAIN 9 1677 VITELLOGENIN.
CC CHAIN 9 71091 LIPOVITELLIN I.
CC CHAIN ?1092 PHOSVITIN.
CC CHAIN ? 1677 LIPOVITELLIN II.
CC DOMAIN 1097 1119 POLY-SER.
CC DOMAIN 1169 1195 POLY-SER.
CC DOMAIN 1215 1218 POLY-ASN.
CC DOMAIN 1219 1329 POLY-SER.
CC DOMAIN 1260 1273 POLY-SER.
CC CARBOHYD 1182 1182 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 1202 1202 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 1217 1217 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 1218 1218 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 1677 AA; 186021 MW; 5F49DA86E43DC35 CRC64;
Query Match 11.8%; Score 84.5; DB 1; Length 1677;
Best Local Similarity 26.1%; Pred. No. 4.8;
Matches 43; Conservative 23; Mismatches 60; Indels 39; Gaps 7;
CC QY 4 SNRSSNQSQDDKQARFGRVRRPWGKFAAEIRDPSPNGARLWLGTFETAEAAAYDRA 63
CC DB 1181 SNSSSSSSSSSSSSSRSHN-HRNTRTLKSKRYQNNSSSSSSSSSEIEIQKPEIF 1239
CC QY 64 AFNLRLHALNFPNEYPRMDYSLRPPYASSSSSSSGSGSTNVSRQNR-----E 116
CC DB 1240 AYFRSHRDLKGFQNK-RGRMS-----SSSSSSSSSSSSQSTLNSKQDAKFLGDSPP 1290
CC QY 117 VPEF-----EYLDKVLK-----ELLDSE--ERKR 139
CC DB 1291 IFAFVARAVRSGLQGYQVAAYTDNRVSRPVQLLATEIEKSR 1335
RESULT 14
ID YMQ5_CABEL STANDARD; PRT; 277 AA.
AC P83351;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein K02D10.5 on chromosome III.
GN K02D10.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2; PubMed=7906398;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey I., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA F Johnston L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U0455; AAA87392.1; -.
CC PIR: T43021; T43021.
CC InterPro: IPR001747; Lipid_transprt_N.
CC InterPro: IPR001846; VWF_D.
CC Pfam: PF01347; Vitellogenin_N; 1.
CC Pfam: PF00094; vwd; 1.
CC SMART; SM00638; LPD_N; 1.
CC SMART; SM00216; VWD_1.
CC Glycoprotein; Phosphorylation; Storage protein; Signal.
CC NON_TER 1 1
CC SIGNAL <1 8 POTENTIAL.
CC CHAIN 9 1677 VITELLOGENIN.
CC CHAIN 9 71091 LIPOVITELLIN I.
CC CHAIN ?1092 PHOSVITIN.
CC CHAIN ? 1677 LIPOVITELLIN II.
CC DOMAIN 1097 1119 POLY-SER.
CC DOMAIN 1169 1195 POLY-SER.
CC DOMAIN 1215 1218 POLY-ASN.
CC DOMAIN 1219 1329 POLY-SER.
CC DOMAIN 1260 1273 POLY-SER.
CC CARBOHYD 1182 1182 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 1202 1202 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 1217 1217 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 1218 1218 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 1677 AA; 186021 MW; 5F49DA86E43DC35 CRC64;
Query Match 11.8%; Score 84.5; DB 1; Length 1677;
Best Local Similarity 26.1%; Pred. No. 4.8;
Matches 43; Conservative 23; Mismatches 60; Indels 39; Gaps 7;
CC QY 4 SNRSSNQSQDDKQARFGRVRRPWGKFAAEIRDPSPNGARLWLGTFETAEAAAYDRA 63
CC DB 1181 SNSSSSSSSSSSSSSRSHN-HRNTRTLKSKRYQNNSSSSSSSSSEIEIQKPEIF 1239
CC QY 64 AFNLRLHALNFPNEYPRMDYSLRPPYASSSSSSSGSGSTNVSRQNR-----E 116
CC DB 1240 AYFRSHRDLKGFQNK-RGRMS-----SSSSSSSSSSSSQSTLNSKQDAKFLGDSPP 1290
CC QY 117 VPEF-----EYLDKVLK-----ELLDSE--ERKR 139
CC DB 1291 IFAFVARAVRSGLQGYQVAAYTDNRVSRPVQLLATEIEKSR 1335
RESULT 14
ID YMQ5_CABEL STANDARD; PRT; 277 AA.
AC P83351;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein K02D10.5 on chromosome III.
GN K02D10.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2; PubMed=7906398;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey I., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA F Johnston L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,

```

RA Wohldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RL elegans";
 RN Nature 368:32-38(1994).
 RP (2)
 RA REVISIONS.
 RA Waterston R.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE SNAP-25 FAMILY.
 CC -!- SIMILARITY: Contains 2 t-SNARE coiled-coil homology domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L14710; AAK84538.1; -;
 DR WormPep; K02D10.5; CE17152.
 DR InterPro; IPR000727; T-SNARE..
 DR SMART; SM00397; t-SNARE; 2.
 DR PROSITE; PS0192; t-SNARE; 2.
 DR Hypothetical protein; Protein transport; Synaptosome; Coiled coil;
 KW Repeat.
 FT DOMAIN 44 106 T-SNARE COILED-COIL HOMOMOLOGY 1.
 FT DOMAIN 179 241 T-SNARE COILED-COIL HOMOMOLOGY 2.
 FT SEQUENCE 277 AA; 31115 MW; EF33EBAED2574B5 CRC64;
 Query Match 11.1%; Score 80; DB 1; Length 277;
 Best Local Similarity 39.7%; Pred. No. 1.5;
 Matches 23; Conservative 10; Mismatches 21; Indels 4; Gaps 2;
 OY 82 PRMDYSLRPYASSSSSSSGSTVSVNQREVFEYFELDDKVLLELDSEERKR 139
 DB 5 PFDDY--RPSAASSTMPVKSYYTTHGYSGEDEADYVERI--EKLQESLDSTER 58
 RESULT 15
 SLAL_YEAST STANDARD; PRT; 1244 AA.
 AC P32790;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cytoskeleton assembly control protein SLAL.
 GN SLAL OR YBL007C OR YBL0321.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RP (1)
 RN SEQUENCE FROM N.A.
 RC STRAIN=DDY 228;
 RC MEDLINE=93328765; PubMed=8335689;
 RX Holtzman D.A., Yang S., Drubin D.G.;
 RA "Synthetic-lethal interactions identify two novel genes, SLAL and
 RT SLA2, that control membrane cytoskeleton assembly in Saccharomyces
 RL cerevisiae.";
 RL J. Cell Biol. 122:635-644(1993).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=93070613; PubMed=1441753;
 RA Delaveau T., Jacq C., Perea J.;
 RT "Sequence of a 12.7 kb segment of yeast chromosome II identifies a
 RP PPR-like gene and several new open reading frames.";
 RL Yeast 8:761-768(1992).
 CC -!- FUNCTION: ESSENTIAL FOR THE PROPER FORMATION OF THE CORTICAL
 CC ACTIN CYTOSKELETON. INVOLVED IN CONTROLLING THE SIZE OF CORTICAL
 CC PATCHES PERHAPS BY REGULATING THE NUCLEATION OF FILAMENTS AT THE
 CC CORTEX.

CC -!- SIMILARITY: SOME, TO SEA URCHIN BINDIN.
 CC -!- SIMILARITY: Contains 3 SH3 domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z22810; CAA80463.1; -;
 DR EMBL; Z35768; CAA84826.1; -;
 DR EMBL; S47695; AAB23985.1; -;
 DR PIR; S25327; S25327.
 DR HSSP; Q01526; 1AWJ.
 DR SGD; S0000103; SLAL.
 DR GO; GO:0005957; C-actin cortical patch (sensu Saccharomycetes); IDA.
 DR GO; GO:0007015; P-actin filament organization; IMP.
 DR GO; GO:0007047; P-cell wall organization and biogenesis; IMP.
 DR GO; GO:0006897; P-endocytosis; IMP.
 DR GO; GO:0007121; P-polar budding; IMP.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00018; SH3; 3.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR ProDom; PD000066; SH3; 3.
 DR SMART; SM00326; SH3; 3.
 DR PROSITE; PS50002; SH3; 3.
 KW Cytoskeleton; Actin-binding; SH3 domain; Repeat.
 FT DOMAIN 8 69 SH3 1.
 FT DOMAIN 70 132 SH3 2.
 FT DOMAIN 353 415 SH3 3.
 FT DOMAIN 868 1205
 FT 16 X 7 AA APPROXIMATE REPEATS OF
 FT T-G-G-A-M-M-P.
 FT REPEAT 868 874 1.
 FT REPEAT 877 883 2.
 FT REPEAT 887 893 3.
 FT REPEAT 923 929 4.
 FT REPEAT 945 951 5.
 FT REPEAT 1003 1009 6.
 FT REPEAT 1020 1026 7.
 FT REPEAT 1031 1037 8.
 FT REPEAT 1048 1054 9.
 FT REPEAT 1065 1071 10.
 FT REPEAT 1084 1090 11.
 FT REPEAT 1129 1135 12.
 FT REPEAT 1155 1161 13.
 FT REPEAT 1170 1176 14.
 FT REPEAT 1185 1191 15.
 FT REPEAT 1200 1206 16.
 SQ SEQUENCE 1244 AA; 135948 MW; 7FD85AA776407624 CRC64;
 Query Match 10.9%; Score 78.5; DB 1; Length 1244;
 Best Local Similarity 25.3%; Pred. No. 13;
 Matches 44; Conservative 28; Mismatches 63; Indels 39; Gaps 9;
 OY 3 SNRNSNN-----QSODDKQARFGVRRPWG-----KPAEIRD-PSRNGAHLWL- 47
 DB 443 SRSRSKSNANASWKDELQNDVVGSAAGKRKSSLSHKKNSATKDFPNKKSRLWVD 502
 OY 48 --GTFETAE-----AARAYDRAAFNLGRHLAILNPPNE---YPRMDYSLRPPYASS 97
 DB 503 RSGTFKVDAEFTCCAGKHLHKANGVKIADKLSNEDLAYVEKLTGFSLE-KFKAND 561
 OY 98 SSSSSSGSTTNVSRNQ-----REVFEYFELDDKVLLELDSEERKR 139
 DB 562 GSSSRGTDGRDSERRRRRLKQEKERDRRLKERELVELKARELLD-EERSR 614
 Search completed: December 17, 2003, 19:21:29
 Job time : 21 secs